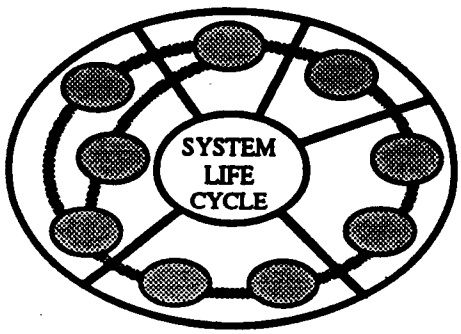
OFFICE OF SOLID WASTE AND EMERGENCY RESPONSE (OSWER)



SYSTEM LIFE CYCLE MANAGEMENT GUIDANCE

Part 3: Practice Paper

Data Management During the System Life Cycle

January, 1989

PRACTICE PAPER FOR DATA MANAGEMENT DURING THE SYSTEM LIFE CYCLE

TABLE OF CONTENTS

EXHIBITS

PRACTICE PAPER FOR DATA MANAGEMENT DURING THE SYSTEM LIFE CYCLE

CHAPTER 1

INTRODUCTION

1.1 Practice Paper Purpose

This practice paper provides details to project managers concerning their responsibilities for data management[1] under the Office of Solid Waste and Emergency Response (OSWER) System Life Cycle Management Guidance. The practice paper describes data management during the system life cycle, and provides guidance concerning major topics that should be addressed by project teams. Data management begins during the Concept Phase, proceeds as requirements are defined and software is implemented, and continues until the application system is terminated or replaced.

This practice paper has three primary purposes:

- Focus each project manager's attention upon ensuring that the data provided by OSWER systems meets program requirements;
- Facilitate successful data management for each project;
- Provide a common approach to data management across OSWER.

This practice paper constitutes a section of Part 3 of the Office of Solid Waste and Emergency Response (OSWER) System Life Cycle Management Guidance.

1.2 Practice Paper Topics

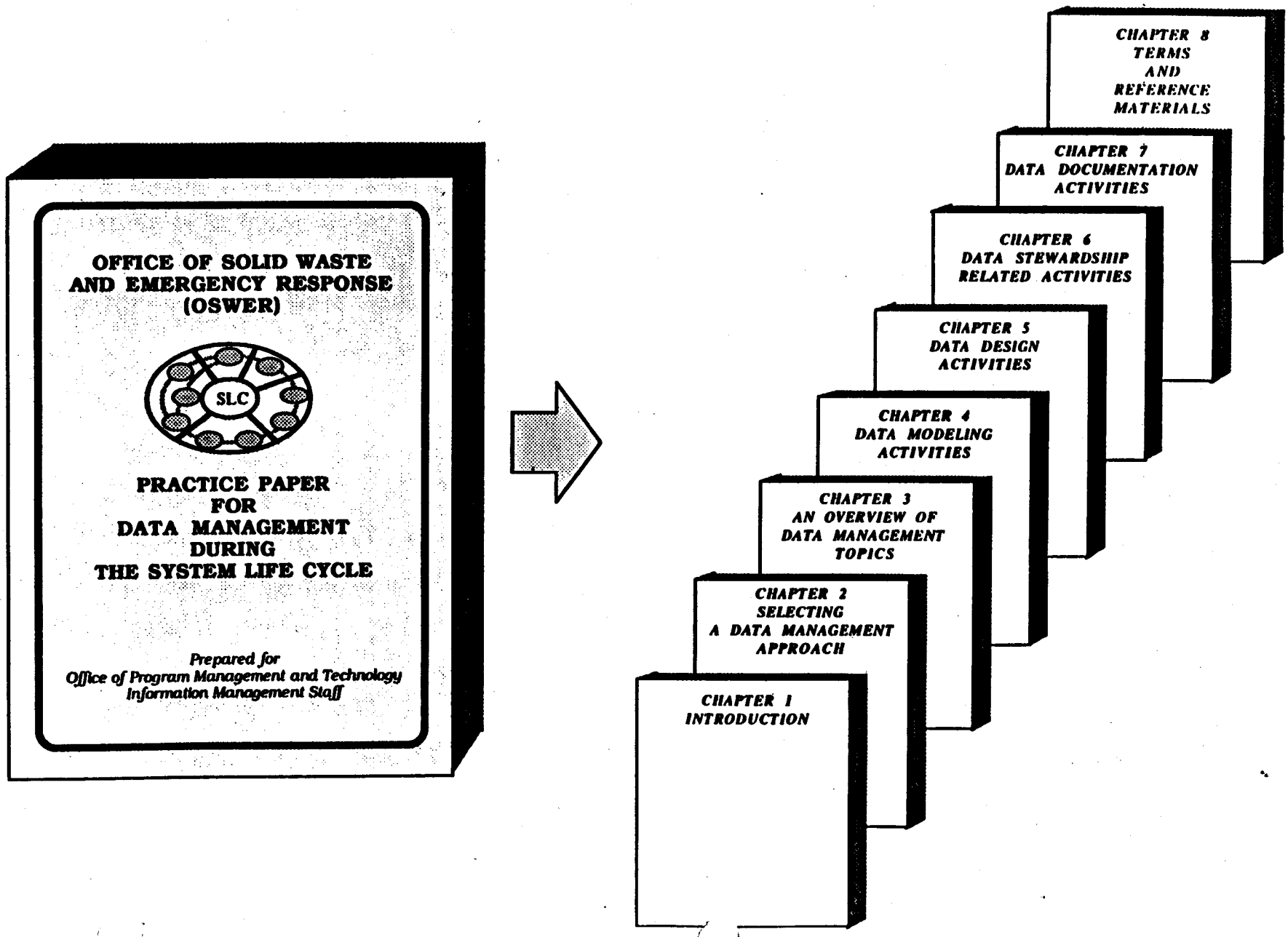
Exhibit 1-1 provides an overview of the structure of this practice paper. Topics addressed in this practice paper include:

- Selection of an approach to data management for a project;
- The details of data management topics applicable to OSWER's Life Cycle Management Guidance;
- The flow of major data management activities, including data modeling, data design, data stewardship, and data documentation.

---

[1]Data management includes data-related activities such as logical data modeling during requirements definition, data base design, data base management, and the documentation of data-related decisions and products.

EXHIBIT 1-1
DATA MANAGEMENT PRACTICE PAPER OVERVIEW



1.3 Data Management Responsibilities

Project managers are responsible for implementing the guidance within this practice paper when OSWER systems and data bases are built. Each project manager will select an appropriate approach to data management for her/his project and ensure that this approach is documented in a Data Management Plan. Project managers will use the Project Management Plan to schedule the activities detailed in the Data Management Plan.

Project managers are also responsible for ensuring that a programmatic requirement for data is identified and documented before data is collected, processed, stored, and distributed. They must work closely with the OSWER program offices affected by a project to ensure that this basic responsibility is met.

OSWER program offices are responsible for supporting the data management process of each project. This responsibility includes providing competent programmatic personnel to identify data requirements and to define the meaning, allowable values, edit criteria, and the level of quality and security of the data. Program offices that assume data stewardship responsibility (see Chapter 6) also determine who will collect data and who will ensure its integrity after it is collected.

1.4 Basic Principles

The data management activities performed during the systems development life cycle are based upon the following basic principles:

- o Data is a valuable resource. Data is collected, stored, and used to support critical OSWER program activities and decisions, making accurate and timely data an important OSWER resource.

- o Data is defined separately from the technology used to collect and store it. OSWER data requirements are recorded clearly prior to designing automated data collection and storage methods, so that program needs are understood and recorded.

- o Accurate information about data is essential. Effective management of data collected by OSWER requires that accurate information about data (metadata) be kept.

- o Common data management guidelines, methods, and tools are used. A common approach to defining, modeling, designing, and documenting data will improve OSWER's data quality and make it easier to share data among systems and offices.

1.5 Why Focus Upon Data Management?

Automated and manual systems provide information to the OSWER program. This information is used, for example, to make decisions affecting public health and safety, environmental quality, and the use of public funds. Without this information, the Office of Solid Waste and Emergency Response could not perform its mission. The data collected, created, stored, processed, and disseminated by OSWER systems are used to create the information OSWER needs to operate.

Since information must be based upon accurate data, future OSWER-funded system development projects will focus upon the provision of accurate and timely data to OSWER programs.

Benefits of increasing the focus upon data management include:

- Ensuring that data collected and disseminated on behalf of OSWER meets programmatic requirements fully, including requirements for accuracy and timeliness;
- Improving management decision-making by providing better access to more accurate and timely data;
- Increasing productivity in the information collection and processing activities of OSWER offices as the understanding and use of available data increases;
- Ensuring that existing data can be shared to the maximum practicable extent, avoiding the cost of redundant data collection and storage;
- Reducing the cost of system maintenance and the time needed to modify implemented systems by designing more stable and flexible data bases.

1.6 How to Use this Practice Paper

Before you read this practice paper, read the OSWER System Life Cycle Management Guidance, including the Data Management Plan exhibits within different sections of the guidance. Then, read this practice paper to learn more information about the topics covered during the system life cycle.

Use this practice paper to guide the development of a data management approach that is appropriate for your project. Then, document your approach in the project's Data Management Plan as the project evolves. Add information to the plan, and modify existing information to reflect the current approach at any point in time.

1.7 Project Participation and Coordination Practice Paper

Read the Project Participation and Coordination Practice Paper before beginning to document your data management approach. The participation and coordination practice paper details who should be involved in project activities, including the activities you will define in the Data Management Plan.

1.8 Configuration Management Practice Paper

This practice paper describes change control of the baselines of requirements, specifications, and operating functionality which are documented in the products of the system life cycle. For instance, the Requirements Data Dictionary is a product which is initially created during the Definition Stage. Changes to this product are controlled using the procedures described in the Configuration Management Practice Paper. Read the Configuration Management Practice Paper before preparing your Data Management Plan.

CHAPTER 2

SELECTING A DATA MANAGEMENT APPROACH

2.1 Introduction

This chapter provides guidance for the project manager who is preparing an approach to managing data-related activities, products, and decisions during the systems life cycle. Your data management approach will have a major impact upon the success of your project and the information provided to OSWER programs.

2.2 What Is a Data Management Approach?

The data-related activities, products, and decisions you decide to address during the system life cycle constitute your data management approach. Your approach also includes the degree of rigor you follow when performing these activities and the level of formality you choose when documenting data-related life cycle products and decisions.

Exhibit 2-1 provides an overview of data management activities during the life cycle. Chapter 3 provides detailed information about data-related activities and products on a stage-by-stage basis.

2.3 Why Choose A Data Management Approach

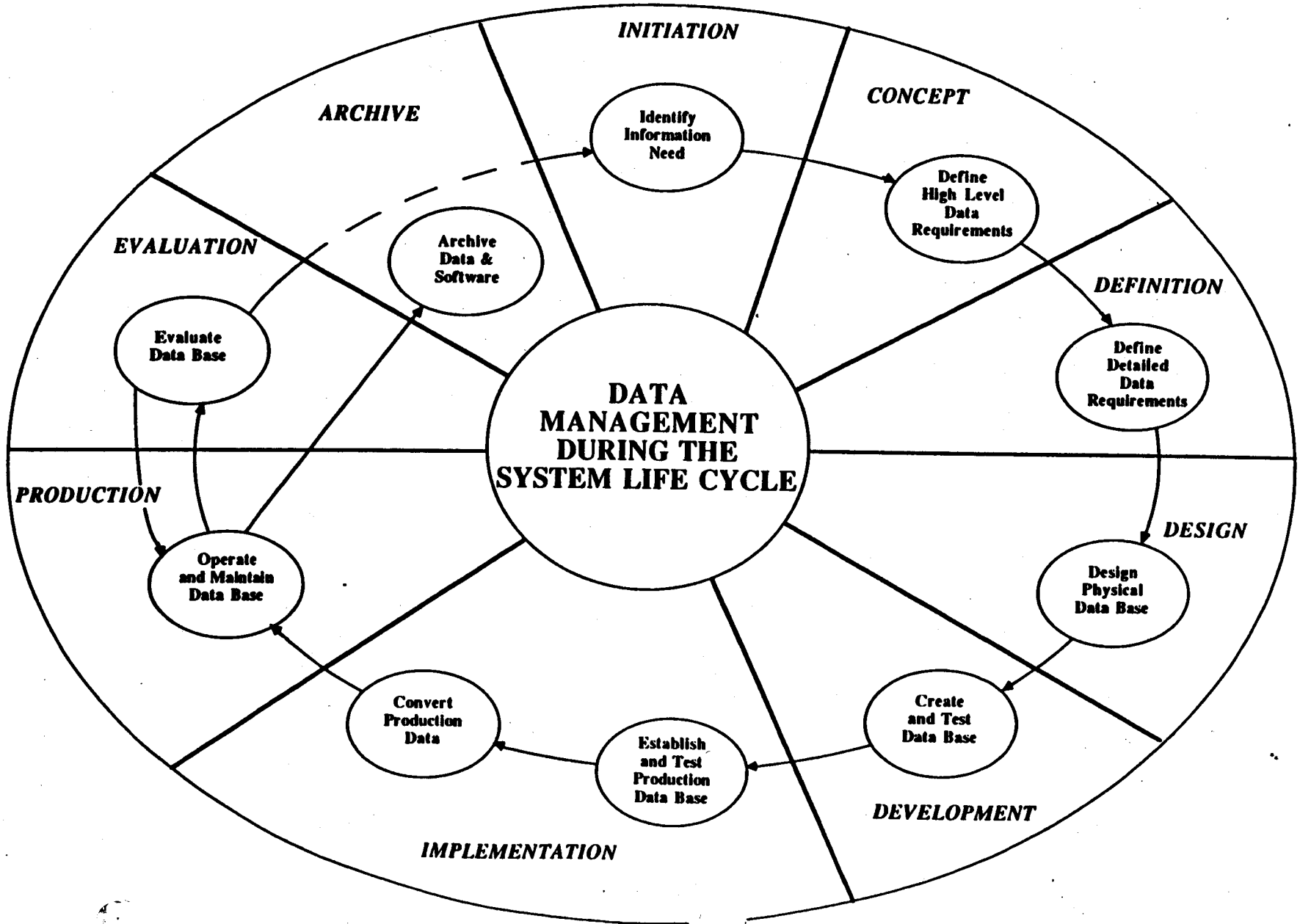
Selecting and implementing a data management approach that is appropriate to your project is a key to your project's success. If you choose an approach that doesn't address data dictionary issues as part of a large, high impact project, you will increase the risk of time and cost overruns for your project, and likely will increase maintenance costs for the completed system and its data. This chapter will give you guidance in determining how much rigor and formalism are required for your project.

One criterion which stands out as a major factor in determining your data management approach is the degree of data sharing. Data sharing includes use of one information system's data by a second system, and using the same data by multiple functions using a single information system. If data sharing is extensive, you should choose a rigorous and formal data management approach. Following this type of approach will minimize unexpected, negative impacts upon your system and the OSWER programs it will support.

2.4 Determining Your Data Management Approach

Adjust your project's data management approach to fit the scope of the system you develop. The degree of data sharing, the scope of organizational impact, and cost of the planned system should be reviewed to determine your data management approach. As the relative degree of each criterion increases, the level of rigor and formalism you select for your data management approach should increase accordingly.

EXHIBIT 2-1
AN OVERVIEW OF DATA MANAGEMENT ACTIVITIES DURING THE SYSTEM LIFE CYCLE



OSWER DIRECTIVE #9029.0C3

2.4.1 Scoping The Project

Begin by determining the scope of data sharing, the organizational impact, and the cost of the project. Extensive data sharing often increases the impact of a system upon the agency. Cost is not as important a criterion in scoping your project, but a large investment in a system should be accompanied by a formal approach to planning and designing the data used to support the system.

Refer to Exhibit 2-2, a chart that will help you scope your project. If any of the statements listed under the high impact column (column two) are true for the project, then your project should be considered high impact.

If your project is not high impact, evaluate each criterion in column one against the information in the medium impact column (column three). If any of the statements under column three apply to your project, then it is a medium impact system. If your project is neither high or medium impact, it is a low impact system for the purpose of determining a data management approach.

2.4.2 Select A Preliminary Data Management Approach

Next, select the data management approach (preliminary) you will use for the project. Refer to Exhibit 2-3 to help select a preliminary data management approach. Choose the decisions, products, and activities you should include in your data management approach, and determine the degree of formalism you will use to document decisions and products. Keep in mind that a primary objective of every system project is to provide accurate and consistent data to the system users.

High Impact Projects

Follow the guidance provided in Chapters 3 through 6 of this practice paper if you have a high impact project. Conceptual data modeling, logical data modeling, and physical data base design are performed. Document the plans, activities, decisions, and products called for in these chapters formally. Data documentation is stored in automated requirements, design, and production data dictionaries. If you are managing a high impact project, appoint a full-time person to coordinate data management activities for the project on your behalf.

Medium Impact Projects

Medium impact projects require subjective judgments concerning the components you select for your data management approach. While conceptual data modeling, logical data modeling, and physical data base design should all be performed, the level of formalism you select for documenting your decisions and products may be less. If you are managing a medium impact project, you may complete data dictionary documentation differently than for a high impact project. Although requirements, design, and production data dictionaries should be kept, only the design and production dictionaries need to be automated.

EXHIBIT 2-2
SCOPING YOUR PROJECT'S IMPACT

CRITERION	HIGH IMPACT	MEDIUM IMPACT	LOW IMPACT
Data Sharing	Data are created, collected, and/or used to support operations of multiple OSWER program offices, Regional offices, and/or state agencies *and* Data are used by multiple EPA offices to provide environmental information to OSWER management, the public, and Congress	Data are created, collected, and/or used by multiple OSWER program offices, Regional offices, and/or state agencies	Data are created, collected, and/or used within a single OSWER program office; no data are created by or collected from Regional offices, and/or state agencies
Organizational Impact	Deemed a national priority system,[1] and is placed on the President's list of national priority systems	Requires FTE support from multiple OSWER program offices, or from Regional offices, and/or from state agencies	Requires FTE support from a single OSWER program office; no FTE support required from other OSWER offices, Regional offices, or from state agencies
Cost		Cost exceeds the available budget support of the sponsoring OSWER office, or the total life cycle cost meets OMB reporting requirements	Sponsoring OSWER office can provide sufficient funding with available budget support, and all costs are below OMB reporting requirements

[1] A national priority system is a system deemed to be critical to the government's mission.
Please note: If any criterion under a column is met, then the project's application is treated as the type of project described in that column.

EXHIBIT 2-3
SELECTING YOUR DATA MANAGEMENT APPROACH

PROJECT SCOPE	DATA MANAGEMENT ACTIVITIES	DOCUMENTATION OF DECISIONS/PRODUCTS	DATA DICTIONARY DOCUMENTATION
HIGH IMPACT	RIGOROUS EXECUTION (See Chapters 3, 4,5,6)	FORMAL (See Chapter 3)	FORMAL (See Chapters 3,7) ● Requirement, Design, and Production data dictionary in automated form.
MEDIUM IMPACT	MODIFIED	MODIFIED	FORMAL ● Requirement, Design, and Production data dictionary in automated form.
LOW IMPACT	MODIFIED	MODIFIED	FORMAL ● Production data dictionary

OSWER DIRECTIVE #9355.0-03

Low Impact Projects

Low impact system projects require a less rigorous approach to activities than the other two classes of projects. Logical data modeling should be done with the resulting data requirements documented. Documentation of production data bases should also be produced and stored manually or in an automated data dictionary. Ensure that the activities and products you select for your approach are appropriate to the scope of your project.

2.4.3 Consult With The OSWER Data Administrator

After selecting a preliminary data management approach, contact the OSWER data administrator in the Office of Policy, Management and Technology for assistance. Review the project's scope and your preliminary data management approach with the data administrator. The data administrator will help you adjust your data management approach to fit the scope of your project and advise you of support capabilities OSWER can provide for your project.

2.4.4 Finalize and Document Data Management Approach

Finalize your data management approach and document it in the project's data management plan. Exhibit 2-4 at the end of this chapter provides the outline of topics for your data management plan. This plan will provide valuable input into project planning and staffing activities, including the preparation of the Project Management Plan.

2.5 Prepare the Data Management Plan

OSWER intends to increase the degree of data integration and sharing to provide improved information management for mission support. Implementation of increased data integration and sharing will be supported by automated software tools for system and data base development, and an OSWER-wide inventory or directory of the data OSWER has collected. Plan to take advantage of these capabilities when you prepare your data management plan.

Record the data management approach in your data management plan early in the Concept Phase. Include those topics described in Chapter 3 which are appropriate. Then, refer to the plan at the beginning of each stage, revising it as you find necessary. See Exhibit 3-1 for a summary of this creation and revision process.

As a project manager, you will find the details in the data management plan helpful when you select staff for your project. Before you prepare your plan, you should read Chapters 4 through 8 of this practice paper to understand the work you will be planning.

Pay particular attention to coordinating activities in four major areas when preparing the substance of your data management plan.

2.5.1 Data Stewardship

Before you can be sure you have identified the relevant data requirements for a project, you must first determine who is authorized to define these requirements. Data stewardship assigns the functional data roles and

responsibilities to the organizations which exercise control over data on behalf of OSWER.

Identify the data steward, definer, and primary user for each high-level data requirement identified during the Concept Phase. Consult with the OSWER data administrator and read Chapter 6 before planning activities related to data stewardship.

2.5.2 Data Modeling

Data modeling activities include identifying data requirements, defining the meaning of the requirements, and structuring the data requirements into logical models. These logical models represent the mission-oriented structure of the data. Data modeling and data models are completely independent of all technology used to store and access data.

When planning data modeling activities, assign an individual with systems analysis skills and OSWER program knowledge to lead the effort. Be sure that "users" of the system and data assume the data definer role described in Chapter 6. It is the user perspective which is critical to the success of data modeling.

You begin data modeling in the Concept Phase and finish the first version of the data model in the Definition Stage. The data model is modified at any point in the life cycle when data requirements change. Since a number of requirements changes will normally occur during the Design and Development stages, you should also anticipate changes to your logical data model during these stages. See Chapter 4 of this document for more information on data modeling.

2.5.3 Physical Data Base Design

Physical data base activities build upon the results of the data modeling activities to produce a physical data base design. This data base design should be done before creating data definition language statements that will be used to create the physical structure of data bases. Data design activities should not begin until after the first version of the data model is complete, since data design activities build upon the results of the data model.

The individual assigned responsibility for data base design must have technical expertise in the data base management system or file structures being used. You should schedule walkthroughs of the design, just as you schedule walkthroughs of your software program designs. The physical data base design walkthroughs should be attended by the data designer, programmers who will use the data structures, the lead data modeling person, and the person responsible for data management.

An initial version of the data base design will normally be required by the end of Design Stage, so that physical data structures (data bases) can be created to support programming and unit testing during Development. See Chapter 5 for more information on data base design activities.

2.5.4 Data Documentation

Collecting, storing, and using data about data (metadata) is crucial to the success of system development project activities. As the individual with primary data management responsibility for your project, you should plan for and manage metadata effectively. Define your data documentation activities early in the Concept Phase, and record them in the Data Management Plan.

You should record metadata describing the meaning, attributes, properties and structures of the data in the logical data model and data base design, as well as data stewardship information. Normally, metadata will be stored in automated software systems called data dictionaries. You must plan to use the metadata of prior stages in subsequent stages whether automated software is used or not. See Chapter 7 for more information about data documentation activities.

EXHIBIT 2-4
DATA MANAGEMENT PLAN OUTLINE

This topical outline contains all the topics that may be contained in the Data Management Plan for a High Impact Project. Taken together, the topics you enter into your Data Management Plan document your project's data management approach.

- Information Need
  - -- Document the Information Need
  - -- Missions Supported
  - -- Scope of the Need
- Data Steward Organizations
  - -- Lead Organization Responsibilities
  - -- Other Organizations Roles
  - -- Data Definers For The Project
- Concept Phase
  - --- Entity List
  - --- Entity Definitions
  - --- Entity Identifiers
  - --- Conceptual Data Model
  - --- Likely Sources of Data
  - --- Information Flow/Data Model Validation
  - --- Data Distribution Plan
  - --- Information Collection Burden
- Definition Stage
  - --- Interview Plans
  - --- Data Analysis By Process
  - --- Entity Normalization
  - --- Conceptual Data Model Revision
  - --- High-Level Data Requirements Revision
  - --- Logical Data Model
  - --- Requirements Data Dictionary
  - --- Data Flow/Logical Data Model Validation
- Design Stage
  - --- Logical Data Model Revision
  - --- Physical Data Base Design
  - --- Design Data Dictionary
- Development Stage
  - --- Data Structures for Programming Support
  - --- Data (structure) Revision Approach
  - --- Data Backup, Logging and Recovery Plans
- Implementation Stage
  - --- Testing Support (See Testing Support Plan)
  - --- Cutover Plans
  - --- Production Data Dictionary

EXHIBIT 2-4
DATA MANAGEMENT PLAN OUTLINE (Continued)

- Production Stage
  - --- Data Base and Metadata Management
  - --- Support of Configuration Management
  - --- Backup, Recovery and Restart
  - --- Role of the Custodian
- Evaluation Stage
  - --- Audit and Evaluation Support Plan
  - --- Response to Evaluation Report
- Archive Stage
  - --- Data Base DDL and Metadata Disposition
  - --- Data Disposition
  - --- Cutover Procedures
- Data Documentation Responsibilities
  - -- Creating Data Documentation
  - -- Maintaining Existing Data Documentation
- Data Quality Assurance Plans
  - -- Responsible Organization
  - -- Milestones and Staffing
  - -- Data Quality Objective Monitoring Plan
- Data Security Requirements and Strategy
  - -- Sensitive Data
- Data Life Cycle Methodologies and Tools
  - -- Metadata Management Approach
  - -- Development & Installation Phase
  - -- Data Management Software
  - -- Operations Phase
- Data Conversion Strategy
- Data Conversion Plan
  - -- Sources
  - -- Media
  - -- Load Programs Required
  - -- Schedule and Staffing
  - -- Validation
- Plan For Physical Flow Of Data
- Data Testing Strategy
- Testing Support
  - -- Kinds of Test Data Bases Required
  - -- Test Data Provision
  - -- Performance Validation Plan
  - -- Responsible Organization
  - -- Projected Testing Support Needed

CHAPTER 3

AN OVERVIEW OF DATA MANAGEMENT TOPICS

3.0 Introduction

This chapter provides additional details of the topics that you can select as a part of your data management approach. If you select one of these topics, refer to this chapter to learn what should be included in your Data Management Plan. Keep in mind that the data management approach and much of the Data Management Plan are prepared during the early part of Concept Phase. In many instances you may not want to enter all of the details of your data management approach in the plan; if not, at least indicate which topics you are addressing.

3.1 Initiation Phase Topic

3.1.1 Information Need

Prepare a brief summary of the information need created by the information management problem discussed in the Initiation Decision Paper. If it is necessary to collect new information or access existing information to solve a problem, then you should select a data management approach and document it in a Data Management Plan. Record the programmatic missions with the information need, and review this information at the start of each phase/stage to be sure it is still valid.

3.2 Concept Phase Topics

3.2.1 Data Steward Organizations

Document the data stewardship responsibilities of the organization sponsoring the project. Then, record the data steward, data collector, data definer, and primary data user for each data entity identified during the Concept and Definition. See Chapter 6 for more information about data stewardship.

3.2.2 High Level Data Requirements

-- Entity List
  - Data Entities[2] about which data is needed (e.g., Employee)

-- Entity Definitions
  - Programmatic definitions of each entity

---

[2]Data entity is a person, place, thing, concept, or event about which OSWER will store data.

EXHIBIT 3-1: DATA MANAGEMENT TOPICS THROUGH THE SYSTEM LIFE CYCLE

PHASE/STAGE / DATA MANAGEMENT PLAN TOPIC	INITIATION	CONCEPT	DEFINITION	DESIGN	DEVELOPMENT	IMPLEMENTATION	PRODUCTION	EVALUATION	ARCHIVE
Information Need Summary	C								
Data Steward Organizations (Plan)		C	R	R	R	R			
High Level Data Requirements -- Conceptual Modeling		C	R						
Data Documentation Plans Responsibilities		C	R	R	R	R	R		
Life Cycle Methodology/Tool Plans		C	R	R	R	R	R		
Physical Data Flow Plan		C	R	R	R	R			
Definition Stage -- Logical Modeling -- Detailed Requirements		C	R						
Data Quality Assurance Plans -- Monitor Data Quality			C	R	R	R	R		
Data Security Requirements -- Strategy			C	R	R	R	R		
Design Stage -- Physical DB Design			C	R					
Data Conversion Strategy				C	R				
Testing Support Plan				C	R	R			
Data Testing Strategy				C	R	R			
Development Stage -- Programming Support				C	R				
Implementation Stage -- Testing Support -- Production Cutover Plans					C	R			
Production Stage -- DB Support Plan -- Production Dictionary Support Plan -- Backup, Recovery, Restart Plans (Procedures) -- Custodial Responsibilities						C	R		
Evaluation Stage -- Support Evaluation -- Response							C	R	
Archive Stage -- Data Base DDL, Metadata, and Data Disposition Plan							C		R

LEGEND:

C = CREATE R = REFINE/REVISE

-- Entity Identifiers (Candidate Keys)
  - Data elements used to uniquely identify each entity (e.g., Employee SSN)

-- Conceptual Data Model
  - Develop a graphic representation of entities and their relationships

-- Likely Sources of Data (by Entity)
  - Identify the likely functions and organizations providing data

-- Validation of the Conceptual Data Model
  - Document your plan to validate that each entity in the high-level information flow is in the conceptual data model. The high-level information flow is done early in the Concept Phase, and is included in the System Concept Document.

-- Data Distribution Plan (by Entity)
  - Record the use of each entity in two 2-dimensional matrices, by function and by distribution/location.

-- Information Collection Burden
  - Plan for any additional information collection burden hours imposed by the solution to the information need.

3.2.3 Data Documentation Responsibilities

Produce and document plans to allocate responsibilities for documenting data requirements, data designs and physical data structures. Look for plans done by earlier projects of the same scope, since this can save you time and effort. Include the data objects to be documented, the attributes to be documented and the functional areas responsible for creating the documentation. Also detail the functional support available to help people create and record metadata, and the media to be used for recording dictionary documentation. Prepare maintenance procedures for dictionary documentation, in coordination with the project's configuration management plan.

3.2.4 Life Cycle Methodologies and Tools

Record the methodologies and automated tools you will use during the project to support data management activities. Be sure to include your explicit plan for managing the flow of metadata (information about data) between methods and tools through the life cycle. These tools might include automated data dictionaries, data modeling and data design tools, and data base management systems.

3.2.5 Plan For Physical Flow Of Data

Record your plans for the physical data (data sets) flowing through the system. This is particularly important when automated data is added to the system from external sources, or when multiple physical data sets need to be updated from a single source of data. Later, timeliness and error correction procedures should be developed.

3.3 Definition Stage Topics

Definition Stage activities and products are especially important, since they define a system's detailed data requirements. Chapters 4 and 7 contain more information about these activities. Topics covered should include:

-- Interview Plans
- Determine and record the name of the interviewees, functional areas to be the subject of interviews, and relationship between the data requirements interview and the functional requirements interview. (You may want to combine these interviews.) Also record the information needed from the interviews in a pre-written interview guide. After the interview, follow-up with a written summary of the results to the interviewee.

-- Data Analysis By Process
- Determine and document your plan to identify and analyze the data elements required by each process during the functional analysis (see SLCM guidance). This analysis plus the logical data model will produce the detailed data requirements for the project.

-- Entity Normalization
- Determine and record the methodology you will use to normalize the entities required by each process. Normalization is the process of reducing a data entity to its most basic form, removing repeating data elements, data elements not dependent upon the key of the entity, and data elements dependent upon the key of other entities. Your plan to perform data analysis should be documented as part of the data analysis topic mentioned above.

-- Conceptual Data Model Revision
- Determine and document your plan for updating the conceptual data model as new entities are determined. The revision should be on-going during the Definition Stage, not held until the end of the stage.

-- Logical Data Model
- Determine and document details of your plan to create logical data models for each functional process, and for the project as a whole. A logical data model is a graphic depiction of the logical, or programmatic, data needed to support an organizational mission. See Chapter 4 for more information. Once you have a draft of your logical data model, hold group review sessions (walkthroughs) to validate it.

-- Requirements Data Dictionary
- Determine and document your plan for recording metadata describing data entities and data elements in the requirements data dictionary. Be specific about responsibilities, activities, and external support you will need.

-- Data Flow/Logical Data Model Validation
- Document your plan to validate the data elements in the logical data model against the data elements in data flows prepared by the system analysts.

3.3.1 Data Quality Assurance

Much of the improvement in data quality that OSWER requires will occur as a result of following this practice paper's approach carefully. However if your system users want to measure the quality of the data in their system periodically, document your data quality assurance plan. If you establish quantitative measures of quality for specific data elements, or combinations of data elements, record your plan for measuring your data bases' ability to meet these measures. Include the following information:

-- Responsibilities
  - What organization is responsible for monitoring data quality
  - Who will monitor or audit data quality
-- Data Quality Objective Monitoring Plan
  - Objectives of the data quality assurance plan
  - What data will be monitored
  - How will it be done and how often
  - How will you resolve problems that are raised
  - Who will monitor problem resolution

3.3.2 Data Security Requirements and Strategy

Include your plan for identifying data security requirements, recording the requirements, and implementing them. Responsibility for identifying sensitive data and protecting the data must be detailed in accordance with the data stewardship roles of the project. Details of data security requirements are recorded in the Security Manual during the Development Stage.

3.4 Design Stage Topics

Create and document your approach to Design Stage data management topics. Topics should include:

3.4.1 Logical Data Model Revision

- Detail your plan for revising the overall logical data model.

3.4.2 Physical Data Base Design (Data Design)

- Document the method you plan to use to prepare a physical data base design (data design) for the project. The physical data base design you prepare will be used to create the physical data structures to support your system. See Chapter 6 for more information on this topic.

3.4.3 Design Data Dictionary

- Enter your plan for recording information (metadata) about the physical data base design in the design data dictionary. Include an outline of what metadata can be copied from the requirements data dictionary.

3.4.4 Role of the Data Custodian

- Identify and document the data custodian for data in the data base. With distributed data bases you will have multiple custodians, so careful planning is needed. Record this information in the design data dictionary during the Design Stage.

3.4.5 Data Conversion Strategy and Plan

If existing data will be used in the new system, include details of your plans for data conversion and the support of data conversion activities. Information to be included are:

-- Sources
-- Media
-- Load Programs Required (automated procedures)
-- Edit Criteria Related to Quality
-- Validation Plan

3.4.6 Data Testing Strategy

Document your general strategy for testing data base performance and functionality to prove you have a viable data base design that meets detailed data and system functional requirements. Most functional test planning information will be stored in the System Test Document.

3.4.7 Testing Support Plan

Record your detailed plans for supporting testing from the Development Stage through the Production Stage. Information includes:

-- Kinds of Physical Test Data Bases Required
-- Test Data Provision
-- Performance Validation Plan
-- Projected Testing Support Needed

3.5 Development Stage Topics

Your approach to Development Stage data management activities should include:

3.5.1 Data Structures for Programming Support
- Record your plan to support software program development and unit testing here. Coordinate your plan carefully to avoid problems. This section will vary depending upon the data base and programming language you use.

3.5.2 Data (structure) Revision Approach
- Document the details of the change control procedures you will use to support changing and adding data structures to the data base during this stage. Refer to the project's Configuration Management Plan for guidance concerning the overall framework you will use. Also document your approach to updating the design data dictionary to keep abreast of all changes.

3.5.3 Data Backup, Logging and Recovery Plans
- Record your plans for backup, logging, and recovery of physical data sets stored in the data structures your project creates. It is suggested that you document your plans for test data bases, and your plans for the production versions of data.

3.5.4 Data Documentation and User Training Materials
Ensure that data management concerns are included in all documentation, not just the data dictionaries, and that user training material includes an emphasis on data creation, collection, validation, and quality assurance issues.

3.6 Implementation Stage Topics

Your approach to Implementation Stage data management topics covered in your plan should include:

3.6.1 Testing (See Acceptance Test Plans)
- Include plans for supporting integration testing and acceptance testing with the data bases that have been built.

3.6.2 Cutover Plans
- Record the activities the data base administration function must carry out to support cutover to the production system. This cutover will usually involve unloading test data, securing the production version of software, and loading data that has been converted from an existing system.

3.7 Production Stage Topics

Topics covered should include:

3.7.1 Data Base and Metadata Management
- Be sure to have procedures and plans for supporting the production data base and keeping the production data dictionary accurate. Also describe procedures to keep the design and production data dictionaries in synchronization when enhancements occur.

3.7.2 Support of Configuration Management
- Coordinate data management with the configuration plans of the project. Refer to the project's Configuration Management Plan for details of the project's approach. Pay particular attention to the updating of data dictionary information, and control of changes requested for data bases.

3.8 Evaluation Stage Topics

Topics covered should include:

3.8.1 Audit and Evaluation Support Plan
- If data management activities are required to support the audit and evaluation process, document the planned activities.

3.8.2 Response to Evaluation Report
- If data management related actions are required in response to points of an evaluation report, document the planned actions.

3.9 Archive Stage Topics

Topics covered should include:

PRACTICE PAPER FOR DATA MANAGEMENT DURING THE SYSTEM LIFE CYCLE

3.9.1 Data Base Data Definition Language and Metadata Disposition
- Document plans to archive or pass to another data base the data definition language statements that create the data structures in the data base, and the metadata in the production data dictionary.

3.9.2 Data Disposition
- Document plans to archive or pass to another data base the data sets (data) in the data base.

3.9.3 Cutover Procedures
- Document plans for the data base administration function to support the cutover from production to archiving or another system.

Chapter 4

DATA MODELING ACTIVITIES

4.1 Introduction

This chapter presents the key concepts of data modeling and lists topics outlined in the Data Management Plan that pertain to data modeling.

4.2 Overview

Data modeling activities relate the data requirements of a project to a programmatic, or end-user perspective. These activities provide essential input into the statement of requirements for a project. There are two levels of data modeling that you will perform during a project:

- o Conceptual Data Modeling
  - a broad look at data requirements

- o Logical Data Modeling
  - expansion of the conceptual model to include detailed requirements

You perform conceptual data modeling during the Concept Phase, and logical data modeling during the Definition Stage. In both cases, the data models are validated by checking their completeness against data flow diagrams prepared by systems analysts, and by reviewing them with end-users. Conceptual data modeling contributes to the preparation of high-level data requirements for the System Concept, and the logical data modeling contributes to the preparation of detailed data requirements that are documented in the requirements data dictionary.

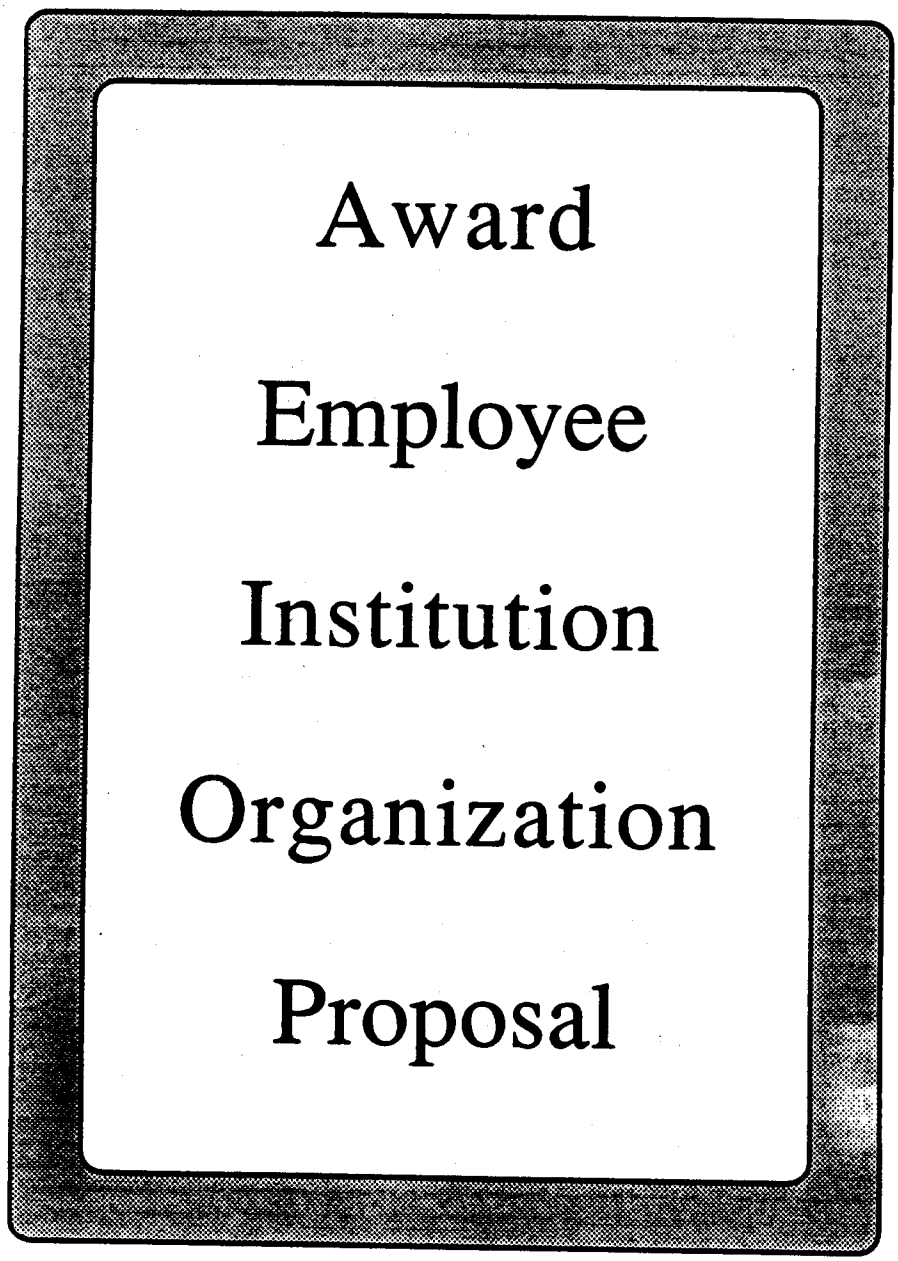
4.3 Conceptual Data Modeling

Conceptual data modeling, the first step in data modeling, begins during the Concept Phase. A conceptual data model shows the kinds of data that should be stored for system users, and the relationships among the data. Often a conceptual data model is used to coordinate the data requirements of one project with the requirements of other projects by providing a framework for assigning data elements to specific, high-level data entities.

The data identified during conceptual data modeling are called data entities. A data entity is a person, place, thing, concept or event about which OSWER needs to obtain data. Facility, Determination, and Permit are all examples of data entities.

You begin conceptual data modeling by preparing a list of data entities for the project. The list can be prepared by examining documents describing the processes to be automated, by interviewing programmatic staff directly, or by working in conjunction with systems analysts during their initial interviews of users. Exhibit 4-1 is an example of a data entity list.

EXHIBIT 4-1: ENTITY LIST EXAMPLE



Part of high level data requirements.

Then, each data entity is defined in programmatic terms, and the data elements that can be used to uniquely identify each occurrence of each entity are identified. Record the data entity definitions carefully, since they are needed for the System Concept document, and for the requirements data dictionary (Definition Stage).

Next prepare a picture of the entities, the relationships between entities, and the data elements that uniquely identify each entity (sometimes called candidate keys). Exhibit 4-2 is an example of such a picture. These pictures are sometimes called entity-relationship diagrams.

Finally, review the information flows, or high-level data flow diagrams prepared by systems analysts to ensure that data entities representing all the data in the data flows and stores on these flow diagrams are represented in the conceptual data model. Additional data entities are defined, added, and the entity-relationship diagram is updated until it is complete.

4.4 Logical Data Modeling

Logical data modeling is a vital activity during the development of any data base, since it includes analysis of programmatic data requirements and provides the starting point for physical data base design. Modeling includes depicting the data required for programmatic functions graphically, and validating the accuracy of the requirement with the users of the data. The logical data model provides a clear, accurate description of data requirements that physical data base designers (see Chapter 5) use to begin design of data bases.

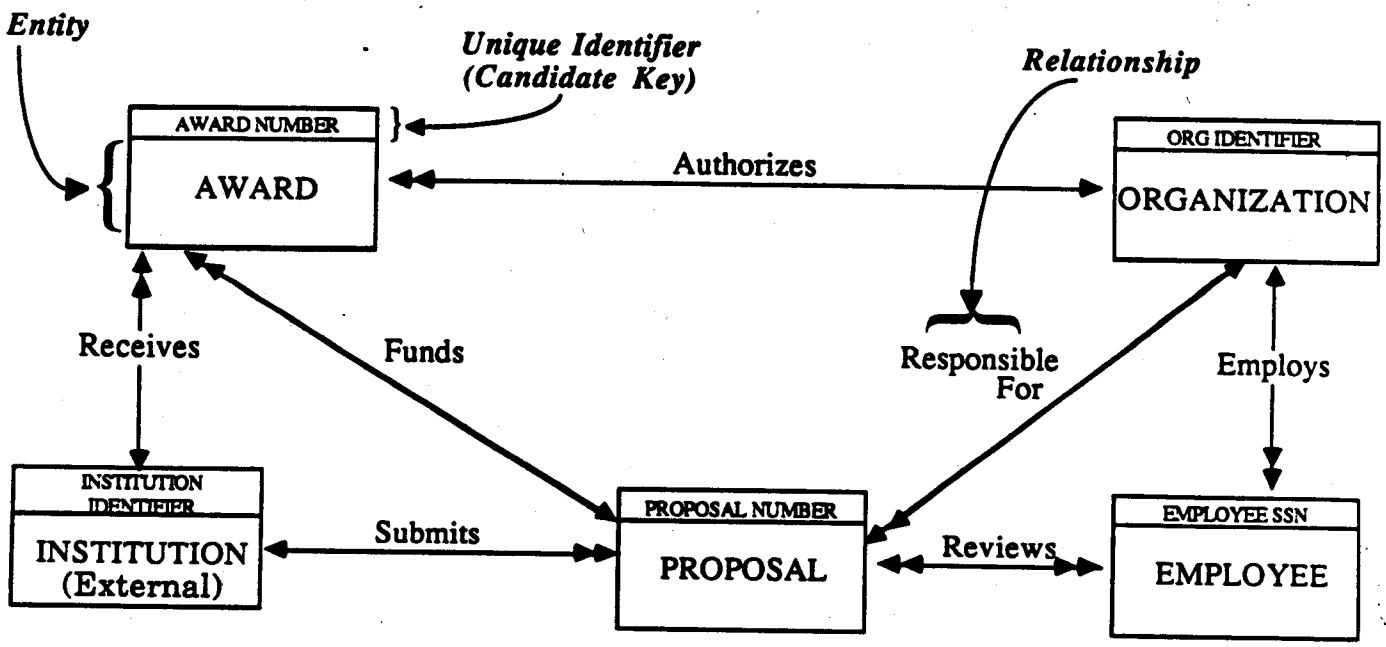
Logical data modeling is performed during the Definition Stage, and the logical model is maintained throughout the life cycle. Logical data modeling builds upon the conceptual data model diagrams and entity definitions that were prepared during the Concept Phase. This model extends the conceptual data model by identifying the data elements required to describe each data entity. While the conceptual data model might contain the data entity "Employee" and the data element "Employee Number" that uniquely identifies the entity, the logical data model would include other data elements, often called attributes, needed to describe "Employee". Examples of such data elements include "Employee Name", "Employee Home Address", and "Employee Birth Date".

Perform a detailed data analysis of the data flows and stores documented by systems analysts during the Definition Stage to expand the conceptual data model into a more detailed, logical data model by adding and defining the data elements required for each entity.

Construct a logical data model diagram that includes data entities and data elements needed by all processes. Then, "normalize" data entities to eliminate logical redundancy, so that your "normalized" logical data model will contain each data element in only one place. See Exhibit 4-3 for an example of a logical data model.

When you normalize, the number of data entities will increase dramatically as you are forced to identify more specific entities that portray sub-types ("Manager", a sub-type of "Employee") and roles ("Assigned Employee") of the entities you had in your conceptual model. Don't be alarmed by this

EXHIBIT 4-2: CONCEPTUAL DATA MODEL EXAMPLE



This conceptual data model describes the relationship between data entities composing the high level data requirements for a proposal award function. Of particular note is the fact that one proposal can result in many awards. Other relationships in the model are:

One	Institution	Submits	Many	Proposals
One	Employee	Reviews	Many	Proposals
One	Organization	Authorizes	Many	Awards
One	Proposal is	Funded by	Many	Awards
One	Institution	Receives	Many	Awards
One	Organization	Employs	Many	Employees

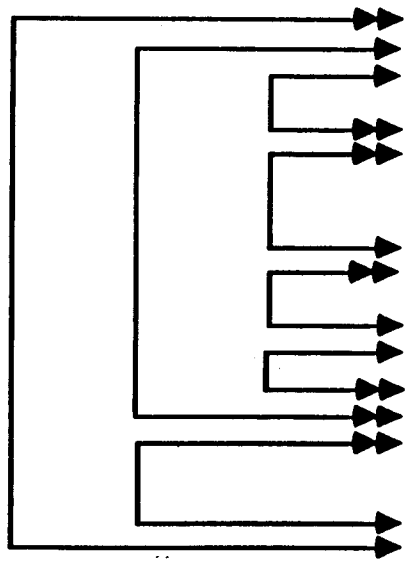
LEGEND

→ Indicates One Entity

→ Indicates Many Entities

EXHIBIT 4-3: LOGICAL DATA MODEL EXAMPLE

PROCESS: PRINT A SUMMARY OF PROPOSALS FUNDED LAST MONTH BY SUBMITTING INSTITUTION



RELATIONSHIP	ENTITY NAME	DATA ELEMENTS (ATTRIBUTES)
	PROPOSAL	PRO #, PRO NAME, PRO SUBMIT DATE, [INS ID]
	PROPOSAL REVIEW	PRO # + EMP SSN # + REV #, REV DETERMINATION, REV START DATE, REV END DATE
	EMPLOYEE	EMP SSN #, EMP NAME, EMP PHONE, [ORG NAME]
	ORGANIZATION	ORG ID, ORG NAME
	AWARD	AWD #, AWD AMT, AWD DATE, [ORG ID], [PRO #], [INS ID]
	INSTITUTION	INS ID, INS NAME

This is a logical data model for the process of printing a summary by submitting institution of funded proposals. The entities here are drawn from a conceptual data model that covers many processes. Once this logical model is normalized, any new entities found are added to the conceptual data model, and used to update the overall logical data model.

LEGEND

+ Indicates Concatenation **Entities in bold** , Separates Data Elements

Primary Keys Underlined [Secondary Keys in Boxes]

proliferation of data entities, it is a sign of your success at "normalizing" entities. This proliferation of entities is caused by your being increasingly more specific in describing and modeling your project's data requirements.

Validate that you have identified all data entities and elements by examining the systems analyst's data flow diagrams again to ensure that all data elements in the data flows and data stores have been accounted for in your logical data model. End-users should then review the model through a formal process, and "sign-off" on the logical model's accuracy.

Finally, add all the descriptions of the data entities and data elements in the logical model, as well as their definitions, to the requirements data dictionary. Contact the Information Management Staff of OPMT to obtain a copy of relevant OSWER documentation standards. Enter data stewardship information into the dictionary (see Chapter 6). The requirements data dictionary and the logical data model diagrams taken together comprise the detailed data requirements for the project.

4.5 Data Management Plan Topics Related to Data Modeling

Topics in the Data Management Plan that cover data modeling are included below. Methodologies and tools should be selected before data modeling begins.

- Concept Development
  - -- Entity List
  - -- Entity Definitions
  - -- Entity Identifiers
  - -- Conceptual Data Model
  - -- Likely Sources of Data
  - -- Information Flow/Data Model Validation
  - -- Data Distribution Plan
  - -- Information Collection Burden
- Definition Stage
  - -- Interview Plans
  - -- Data Analysis By Process
  - -- Entity Normalization
  - -- Conceptual Data Model Revision
  - -- High-Level Data Requirements Revision
  - -- Logical Data Model
  - -- Data Flow/Logical Data Model Validation
- Design Stage
  - -- Logical Data Model Revision
- Data Security Requirements and Strategy
  - -- Sensitive Data

  (Identified During Data Analysis and Modeling)
- Life Cycle Data Management Methodologies and Tools

Chapter 5

PHYSICAL DATA BASE DESIGN ACTIVITIES

5.1 Introduction

This chapter presents an overview of physical data base design activities and lists the topics in the Data Management Plan that pertain to physical data base design. Physical data base design activities support the storage of data needed by a system's users. Contact the National Data Processing Division (NDPD) for a copy of the procedures for implementing data bases on NDPD equipment.

5.2 Overview

Data design activities have been a part of system development practices for some years now. The product of these activities has been data structures, for example physical record types, that support a system's software programs. The development of physical data base structures without data modeling has resulted in unstable designs that increase maintenance costs, and limit the usefulness of data stored in these structures. So don't begin your data management activities by doing a physical data base design.

Physical data base design activities begin during the Design Stage after the logical data model is completed. Information contained in the logical data model and the requirements data dictionary is used as a starting point for the design process. The designer transforms the logical data model into an initial physical data base design that can be implemented in a data base management system. Data definition language statements are produced during the Development Stage to provide common data structures for programming and unit testing support. During the Development and Implementation Stages, the design is altered to meet the performance requirements of the system, and the data definition language statements are also revised.

5.3 The Physical Data Base Design Process

The physical data base design process must include participation by someone with an expert knowledge of the DBMS being used. You may decide to do physical data base design in a single step, but if you have a large and complex system, you may want to plan to follow the steps outlined below. There are three significant products in the physical data base design process that you may produce to ensure thoroughness:

- o Data driven physical model
- o Process driven physical model
- o Location driven physical model

These three products are the focal point of the process outlined below. Please note that in some circumstances you may combine two or three of the models and produce a single model instead.

5.3.1 Step 1: Data Driven Physical Model

Begin your physical data base design process by taking your logical data model and transforming it to fit the structure of the data base management system (DBMS) you are using. If you are using a DBMS that can implement a relational data model directly, then this will not require much work at all, since your normalized logical data model will be structured in relational terms. With a relational DBMS the logical model is the data driven physical model. If you use a relational DBMS and have no requirements for high-volume processing or distributed processing, your physical data base design is completed.

If you are using a DBMS such as FOCUS or SYSTEM 2000 that does not support the relational data model, you will have to transform your logical data model into the hierarchical or network structures which these other DBMS' support. This activity involves allocating data elements from your logical data model to more than one physical record or segment, depending upon the DBMS being used. Assign an individual with expert-level knowledge of the DBMS package being used to perform this task. Ask the DBMS vendor if they can provide design examples or material for you.

5.3.2 Step 2: Process Driven Physical Model

Review the transaction volumes and service requirements of particular software programs to determine if high volume or very fast response time is required for some transactions. If so, alter your data driven physical model to place the data elements needed by the high volume or high service requirement programs together in one physical record. This introduces physical redundancy that may be justified by your application. Schedule a walkthrough of the process driven physical model with systems analysts and programmers, the data base designer, and the project's data administration staff.

5.3.3 Step 3: Location Driven Physical Model

If you plan to distribute your data base(s) at more than one location and run software against the distributed data, then you will have to modify your process driven physical model to account for the replication of data needed to support distributed processing. This is not a trivial activity. Make certain that the design team has expertise in the design of distributed data bases using the selected DBMS package.

5.4 Using Your Physical Data Base Designs

Normally, the process driven physical model will be the initial physical data base design for your project. If you are using a distributed data base for your project, you will have to complete a location driven physical model and use it for your physical data base design. Whichever is the case, use the initial physical data base design to support the preparation of data definition language statements for your DBMS during Development. If no DBMS is used, these statements will be replaced by copybooks or copy libraries that can be shared by programmers.

Once the physical data base design is done, create the design data dictionary (if you haven't already done so), by copying the definitions for data elements in the requirements data dictionary into the design data dictionary and

adding technical information that describes your physical data base design. Maintain the design data dictionary and transfer it to the production data dictionary at Implementation Stage. See Chapter 7 for more details on documentation.

The physical data model (physical data base design) is created during the Design Stage, and the data definition language statements (or copybooks) are produced early in the Development Stage to support development and unit testing of programs. These data structures are maintained throughout the rest of the system life cycle under the control of configuration management.

5.5 Data Management Plan Topics Relating to Physical Data Base Design

Topics in the Data Management Plan that cover physical data base design are included below.

Design Stage
-- Logical Data Model Revision
-- Physical Data Base Design
-- Design Data Dictionary

Development Stage
-- Data Structures for Programming Support
-- Data Structure Revision Approach

Implementation Stage
-- Testing Support

Life Cycle Methodologies and Tools
(Supports physical data base design activities)

Data Security Strategy
(Requirements are supported by the physical data base design)

Plan for Physical Flow of Data
(Supported by the physical data base design)

Testing Support Plan
(Supported by the physical data base design and DBMS)
-- Performance Validation Plan

CHAPTER 6

DATA STEWARDSHIP RELATED ACTIVITIES

6.1 Introduction

If you are the individual responsible for data management, you should create a plan to determine and document data stewardship for the data your system will collect. This plan should be done early in the Concept Phase, since several concept activities implement data stewardship.

6.2 Background

The Office of Solid Waste and Emergency Response uses thousands of data elements while performing the missions assigned to it. The flow of OSWER program data is often lengthy and complex, as data is collected at the state and regional levels, then forwarded to OSWER for analysis, with the results being disseminated to Congress, other agencies, the states, and regions. Managing the complex activities, responsibilities and relationships which arise from these data flows requires a method of determining which organizations involved in the data flows are responsible for which data-related activities.

6.3 Guidance Objectives

The data stewardship guidance presented in this chapter has the following objectives:

a. Facilitate the assignment of responsibilities for data definition, collection, processing, storage, and use, when systems and data bases are being built.

b. Ensure data meets mission requirements by assigning accountability for data stewardship.

c. Ensure that data definition, collection, processing, and storage methods within OSWER systems conform to applicable guidance.

d. Facilitate data sharing and reuse by clarifying roles and responsibilities involved in the definition, collection, processing, storage, and use of data.

6.4 What Is Data Stewardship?

Within OSWER, organizations may share data, but do not individually own it. OSWER defines data stewardship as the functions and responsibilities of an organization that exercises programmatic control over data on behalf of OSWER. Organizations that require data to be collected, processed, stored or used in support of OSWER's mission have stewardship responsibilities. These responsibilities include ensuring that:

1) Data is clearly defined and documented in compliance with established directives.

2) Data that is collected is of sufficient quality to support OSWER's missions.

3) Only data relevant to OSWER's missions is collected.

4) Data is reused wherever appropriate within OSWER.

5) Data management practices for data under the organization's stewardship conform to OSWER guidance.

6.5 Which Organization Is A Data Steward?

The data steward role is separate and distinct from the role of a project manager. Data stewardship assignment parallels OSWER program management responsibility. The same organization with overall responsibility for OSWER program mission performance is responsible for ensuring that the quality data required to support the mission is defined, collected, processed, stored and presented in a timely and cost-effective manner. If a project involves support of more than one program, then multiple data stewards may be involved for the data elements used by the project's system. Exhibit 6-1 provides an overview of the data steward role.

6.6 What Are The Rights and Responsibilities of a Data Steward?

Data Steward Rights. The data steward organization appoints an individual to execute its rights within the data stewardship guidance. The data steward individual sanctions or approves the definition, collection, processing, storage and use of data under his/her stewardship. **This individual also has the right to designate the data definer, the data collector and the custodian (see section 6.7).**

Data Steward Responsibilities. The data steward organization is responsible for ensuring that all the functional data roles are performed adequately to provide data of sufficient quality to support OSWER program missions. Other responsibilities include ensuring that only data relevant to OSWER missions is collected, that data is reused wherever appropriate, and that data is clearly defined and documented in conformance with EPA and OSWER guidance. The steward organization appoints an individual to execute its responsibilities within the data stewardship guidance.

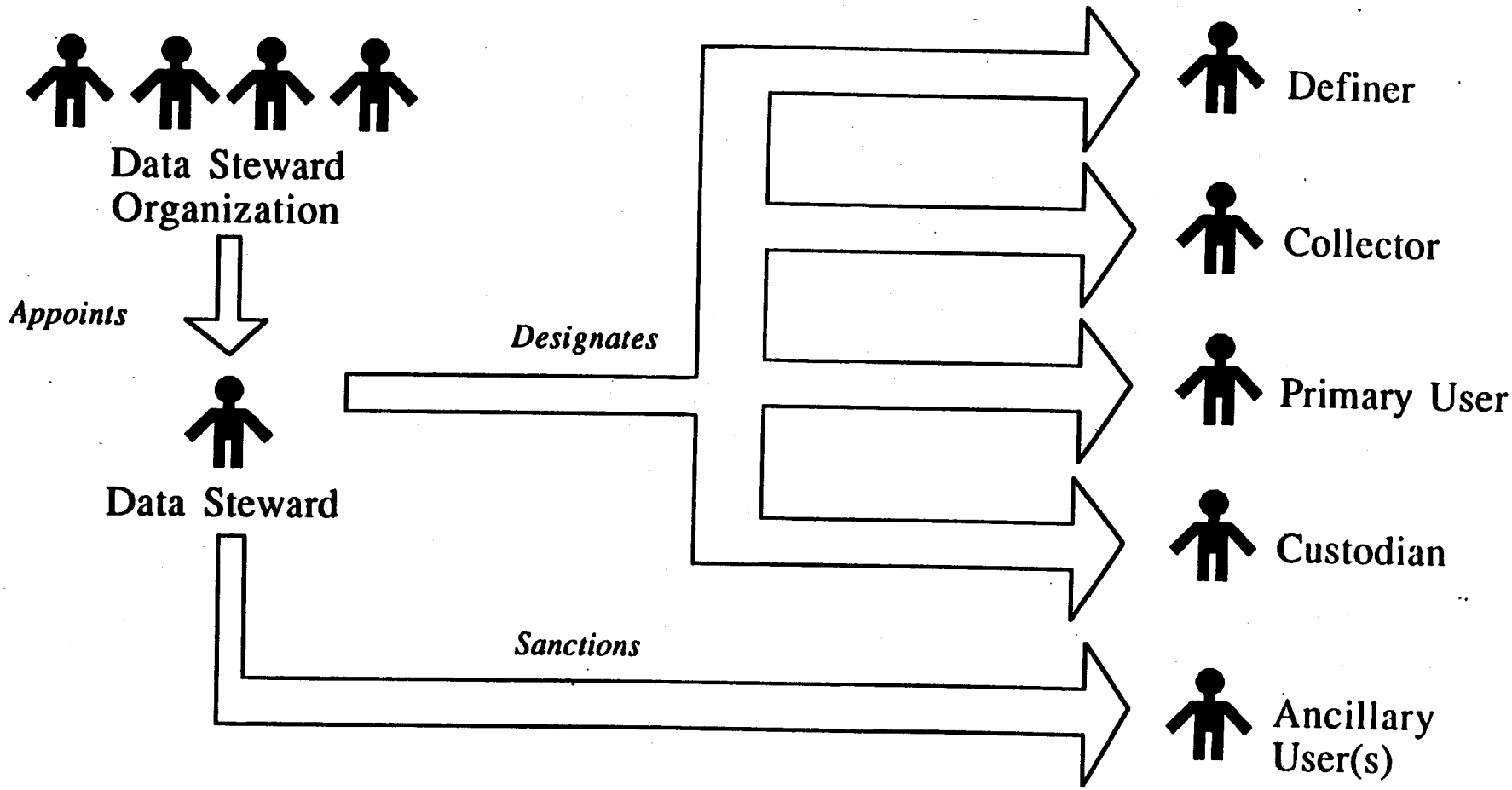
6.7 What Are Functional Data Roles?

Exhibit 6-2 summarizes the function of each data role. A longer discussion of each functional data role follows.

a. Definer.

The organization or function responsible for determining and documenting the name, description, and other attributes of data required to support an OSWER mission is the data definer. The data definer prescribes the derivation rules and the formats to be used for data derived from other data

EXHIBIT 6-1: THE DATA STEWARD ROLE



Rights: Sanction definition, collection, processing, storage, and use of data. Designate data definer, collector, primary user, and custodian.

Responsibilities: Ensures functional data roles are performed adequately to provide data of sufficient quality to support OSWER program missions.

EXHIBIT 6-2: FUNCTIONAL DATA ROLES

ROLE	***FUNCTION***
 Definer	Defines data requirements, meaning, content, and documents the needed data in compliance with OSWER guidance.
 Collector	Collects data required by OSWER missions.
 Primary User	Uses data to perform mission functions. The primary data user is often the data definer, as well. Supports testing during the systems life cycle.
 Custodian	Responsible for supporting the storage and processing of data. Ensures the physical integrity of data used to support OSWER missions while providing access to the users. Implements changes.
 Ancillary User(s)	Uses data defined and stored to support the missions of others. The Ancillary user requests appropriate approval from the Data Steward for accessing and using the data sought.

elements. The data definer may or may not be the same person or organization as the collector. Usually, the data steward designates the primary user(s) of data designated as the data definer. When multiple functions or organizations use the same data to support important OSWER program functions, a joint data definition effort will be organized. The data definer is responsible for ensuring that OSWER and EPA data administration standards are met when defining and documenting data.

b. Collector.

The organizations or functions responsible for collecting the values of data required by an OSWER program are called the data collectors. The data steward organization designates data collectors to direct and manage collection of data for an OSWER program, after consulting with the collector organization to ensure it can assume this role. The data collectors may not terminate the collection of data without the authorization of the data steward organization.

c. Primary User(s).

The organization or function with the most important requirement to collect, store and process data to perform a current or future OSWER mission function is the primary user. Usually, the data steward organization designates the primary user as the definer of the data the organization requires. Primary user organizations also support user testing of systems during the life cycle. If it is impossible to select a single primary user organization from among several users of the same data, then a joint data definition effort will probably result. The primary user organization is often, although not always, the data steward organization.

d. Custodian.

The organization responsible for storage and processing of data is the data custodian. The functions carried out by the custodian include those which have traditionally been performed by ADP organizations. That is, the custodian has physical custody or direct control of the data, software and other ADP components used to store, process, communicate and present data. The custodial role ensures the physical integrity of data, and safeguarding the storage media. During the life cycle, the custodial role may be assigned to the project team until the data base (and system) is turned over to the operating organization during the Implementation Stage. As small computers have become more common, it is not unusual for custodial responsibilities to be split between different organizations, or to be assumed directly by non-ADP professionals within the primary user organization.

e. Ancillary User(s)

The ancillary users use data to perform OSWER functions, and report results to management, the Congress, and to others outside the agency. Unlike the primary users, ancillary users must rely upon others to define and allow them access to the data.

6.8 Implementing Data Stewardship

During the Concept Phase, you should contact the OSWER Data Administrator and discuss the scope of your project's application and the general types of information you anticipate the application will need. If you and the OSWER data administrator determine that multiple data steward organizations will be involved in a project, plan to involve all of these organizations beginning in the Concept Phase. Since implementation of stewardship at the data base or project level is not feasible when multiple data steward organizations are involved, record the steward organization for each data entity you identify during Concept Phase. While there may be multiple data stewards for the data collected by a system, there can be only one data steward per element of data. Enter this stewardship information in the requirements data dictionary during the Definition Stage. Ensure that data definers, primary users, collectors, and custodians are recorded for all data elements.

As noted in Exhibit 6-3, when one organization defines, collects, and uses a file, that organization is the probable data steward. However, when you find data definition, collection, and use split between multiple organizations you face a more difficult problem when determining stewardship. Line 2 of Exhibit 6-3 provides an example of multiple organizations collecting and using the same data file. When this occurs either of the primary user organizations may be selected as the data steward, but not both. Data stewardship of data cannot be shared between organizations. A single data steward organization must be selected for clear accountability.

If your project is developing a high-impact system, you are likely to find that several data steward organizations will be involved. For instance, the application may collect enforcement data at the request of OWPE, and clean-up data at the request of OERR. Be sure that both offices understand their stewardship responsibilities, and work with them to provide the data they require.

Use data dictionaries to save time when recording stewardship information. For instance, if one steward organization sponsors an entire data base, then record the stewardship information as part of the data base dictionary documentation, and use the dictionary reporting facilities to link this information to all data elements in the data base.

Data Stewardship Implementation Activities

Step 1 - During the Concept Phase identify the organizations that will likely be the data steward organizations for the data your project will require.

Step 2 - Have the data steward organization appoint a data steward individual to assume data stewardship responsibilities.

Step 3 - Work with the person appointed to determine the primary data user and assign a data definer or definers for the project. Consult with the OSWER data administrator. Involve the data definer(s) in identifying data entities as your project team performs conceptual data modeling.

Step 4 - After you have completed your data entity list (see Chapter 4), have the data steward assign specific responsibility for defining each data entity's meaning in mission terms. This same data definer will also define data elements to describe the data entity more fully when the Definition Stage is reached. Record the name of the data definer and data steward organization for each data entity in the System Concept and in the definitions of data entities you create.

Step 5 - During the Definition Stage, record the data steward and the data definer organizations for each data entity in the requirements data dictionary and the detailed data requirements. Design your data dictionary entries so that dictionary users can learn the data steward and definer for each data element, too.

Step 6 - During the Definition Stage, determine the function and organization to collect values for each data element; record the data collector in the requirements data dictionary.

Step 7 - During the Design Stage, have the data steward name the organization that will support technical operation of the system and data base once it is in a production mode, and designate that function or organization as the custodian for all data elements in the data base. Record this information in the design data dictionary along with the steward and definer information you added to the requirements dictionary earlier.

EXHIBIT 6-3
DETERMINING DATA STEWARDSHIP

REVIEW FUNCTIONAL DATA ROLES TO DETERMINE PROBABLE DATA STEWARDS

DATA OBJECT	DATA DEFINER	DATA COLLECTOR	PRIMARY USERS(S)	NON-PRIMARY USER(S)	CUSTODIAN	PROBABLE DATA STEWARD
File 1	ORG A	ORG A	ORG A	ORG none	ORG C	ORG A
File 2	ORG A,B	ORG D	ORG A,B	ORG D,E	ORG C	ORG A or B (not both)
Data Element Y	ORG B	ORG D	ORG B	ORG A,D	ORG D	ORG B
Data Element Z	ORG D	ORG D	ORG D	ORG A	ORG C	ORG D
Data Base X	ORG A	ORG A	ORG A	ORG B,D	ORG C	ORG A

ORG A = OSWER program office A
ORG B = OSWER program office B
ORG C = EPA data processing organization C
ORG D = EPA regional office D

CHAPTER 7

DATA DOCUMENTATION ACTIVITIES

7.1 Introduction

This chapter describes the essential data documentation activities that must be performed during the system life cycle. While documentation is required of other life cycle products, such as the System Concept, this paper covers only the essential documentation of data requirements, physical data base designs, and production data base structures. The documentation must be completed to support later phases of the life cycle, reduce maintenance costs, and provide an audit trail from requirements to production data bases. Careful management of this documentation will allow you to track your project's data requirements through to the production data base.

7.2 Data Documentation Products

The OSWER Systems Development Life Cycle Management Guidance specifies that project teams document detailed data requirements, physical data base designs, and production data bases in data dictionaries. Each of these three data dictionaries serve different purposes. While paper or a single software system can hold all three of these dictionaries, the dictionaries should be created discretely.

Dictionary Type	Role
Requirements Dictionary	Contains High Level and Detailed Data Requirements
Design Dictionary	Contains Physical Data Base Design
Production Dictionary	Contains Production Data Base Design

7.3 Requirements Data Dictionary

The requirements data dictionary is produced by data modeling activities that take place during the Concept and Definition Phases of the project. Metadata recorded about data entities and data elements include their names, programmatic definitions, purpose in the OSWER programs, data steward, data definer, and source. Precise guidance regarding the metadata to insert into the requirements data dictionary can be obtained by contacting the Office of Program Management and Technology Information Management Staff.

Many automated software development tools provide good support for the requirements data dictionary. If automated tools have been used to perform data modeling, then some of the documentation in the requirements data dictionary can be loaded from the software development tool into the data dictionary. Likewise, automated tools used to support physical data base design should

contain metadata derived from your data modeling tool, or from the requirements data dictionary. If the metadata used in each tool is based upon the metadata produced by the preceding tool, you will save time and effort while minimizing the risk of inaccurate metadata being used to produce a data base design.

7.4 Design Data Dictionary

The design data dictionary is initially created in the Design Stage and is maintained through the rest of the life cycle. This dictionary contains detailed data documentation of the physical data base design (see Chapter 5). It also contains programmatic and administrative information stored in the requirements data dictionary earlier. Programmatic definitions of data elements stored in the requirements data dictionary are entered into this data dictionary, as the physical design information is entered. If software is used to store and use metadata, then the same piece of software can sometimes be used for both dictionaries. This is a project decision.

Technical information about the physical data base design is entered into the design data dictionary. It describes the design of the physical data base structures and the manner in which these structures are implemented in the test versions of the data base. For this reason the design data dictionary contains many more types of metadata than the requirements data dictionary. In addition to data elements, the design dictionary contains documentation of data bases, physical records, segments, data sets (or files), and keys. Information such as block sizes, data set allocation, and physical size limits are documented in this dictionary.

If your project is using a data base management system which is controlled by an active data dictionary, then use the active data dictionary for the design data dictionary. These active dictionaries contain only the metadata for the data bases they control, and they can save you time and money during system maintenance.

7.5 Production Data Dictionary

The production data dictionary is initially created in the Implementation Stage to support system integration and acceptance testing. Metadata from the last version of the design data dictionary is moved to the production data dictionary at the beginning of the Implementation Stage. Exactly the same metadata is in the last version of the design dictionary and the initial version of the production dictionary. The only difference between them is that the former describes a data base under development and the latter a data base in the final steps of testing and implementation.

You should continue to maintain a design data dictionary to support testing of new versions of the data base throughout the rest of the life cycle. Once the production dictionary is created, the design dictionary contains the same information as the production dictionary, plus proposed changes. If you use a data base management system that is controlled by an active data dictionary, use the active DBMS dictionary as the production data dictionary. The "move" of metadata from the design to production dictionary can usually be done by simply changing the status indicator on those dictionary objects being moved into production.

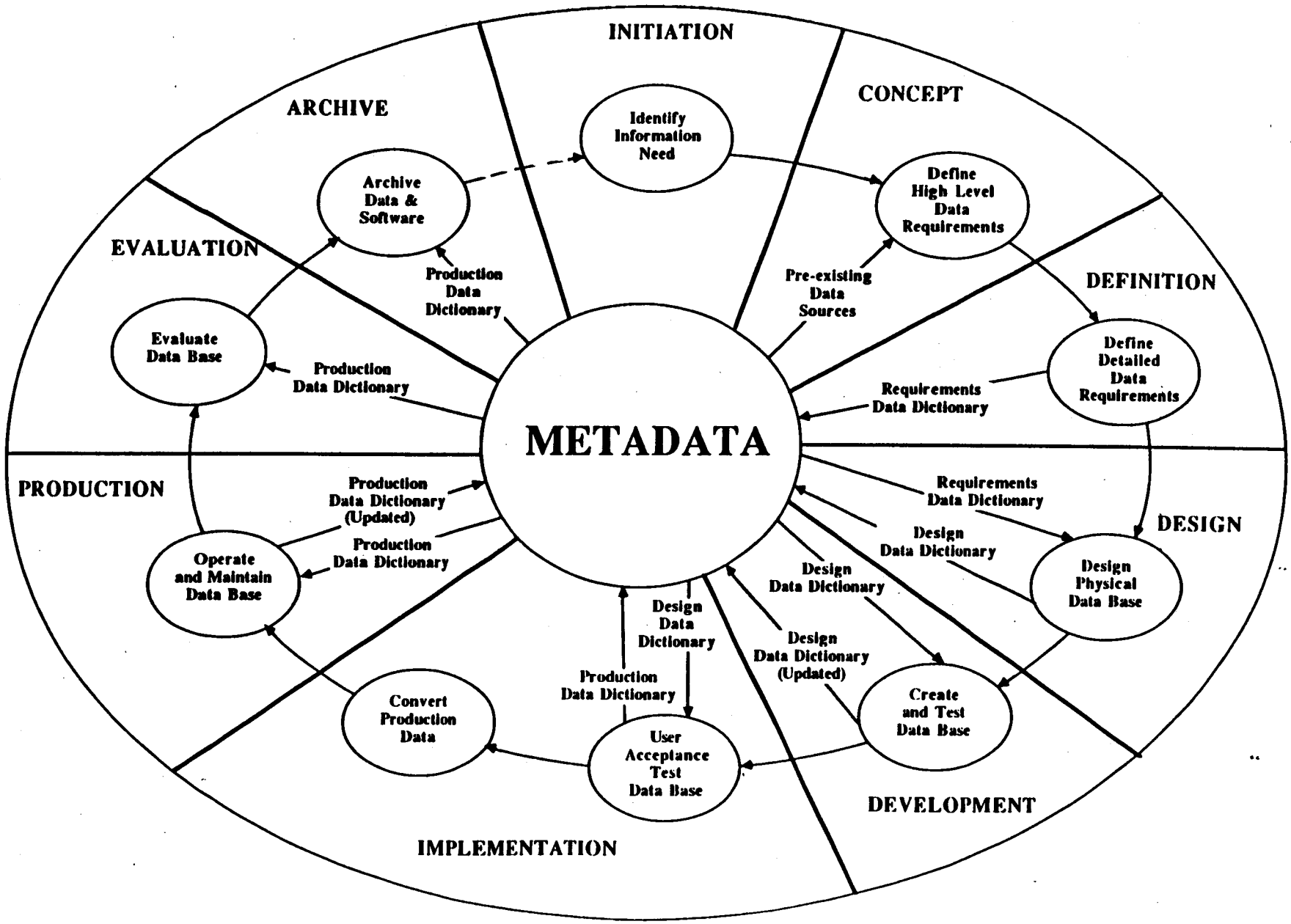
Once metadata is moved into the production data dictionary it is secured against changes or deletion. Entries in the production data dictionary are never written-over or deleted, but new versions are created that reflect any changes. Old versions are archived to create a perpetual history of the structure of the production data base. The production data dictionary is maintained up to the point when the data base is no longer used. Once the data base is slated for conversion to another data base, or archiving, the metadata in the production data dictionary is archived along with the production data and software programs.

7.6 Importance of Metadata Management

As technology has evolved, data dictionaries have taken on an increasingly important role. Many of the most modern, relational data base management systems require that a production data dictionary be loaded before the DBMS is used. Now, system and data base design activities are being supported by automated software tools that are also designed around data dictionaries of their own. As the individual responsible for data management for your project, you will have to plan and monitor your project's collection, use, and transfer of metadata throughout the system life cycle. If you fail in this task, you could delay the project, increase the cost of the project, or cause a data base to be implemented that doesn't meet OSWER requirements. You would be wise to consult with the OSWER data administrator concerning the availability of metadata management software, including the availability of an OSWER-wide data resource inventory, to support your project.

Exhibit 7-1 illustrates the use of data dictionaries during the system life cycle.

EXHIBIT
DATA DICTIONARIES DURING T... SYSTEMS LIFE CYCLE



OSWER DIRECTIVE #9026.02a

CHAPTER 8

TERMS AND REFERENCE MATERIALS

8.1 Glossary of Terms

This chapter defines key terms used in OSWER's System Life Cycle Management Guidance and this practice paper.

TERM	DEFINITION
Access	The operation of viewing or copying (including extracting) data.
Approval	An examination of life cycle products, and the results of the project review process, by OSWER program management. The approval process has three purposes: first, to confirm the results (i.e., the concepts, products, and management direction) of life cycle efforts to date; second, to approve continuation with the next stage of the life cycle; and third, to confirm the continued commitment of resources to the project. The OSWER life cycle model requires formal approvals at the end of the Initiation and Concept phases, and the Definition, Design, Development, and Implementation stages.
Archive	The third stage of the Operation phase, and the final stage of the system life cycle. Its purpose is to terminate the operation of the system in an orderly, planned manner, ensuring that software and data are properly archived or incorporated into other systems.
Audit	A standard-oriented examination of the products and related documentation contained in a baseline to assure that they are complete, clearly presented, and internally consistent. The OSWER life cycle provides for five audits; Concept, Definition, Design, Development, and Operational. Any audit may be repeated as necessary.
Baseline	The set of life cycle products and other related documentation which officially comprise the system at a given point in time. The OSWER life cycle model provides for five baselines: Initiation, Definition, Design, Development, and Operational. The products contained in each baseline are always reviewed prior to inclusion in the baseline.
Change	An alteration to the system or data base(s) for maintenance or performance purposes, without affecting the functionality or structure of the system or data base(s).

Change Control	A process for controlling modifications to a system. Change control provides a review of requested modifications, and consideration of their impact on a system, before they are made; it also ensures that changes are made in a manner that does not disrupt ongoing system operation.
Concept	The second phase of the system life cycle. This phase provides a high level of functional and data requirements that relate to an information management problem, and a comprehensive model of a solution to meet the requirements.
Conceptual Data Model	A depiction of data requirements from an organizational perspective. Corresponds to the conceptual schema of a three schema architecture as defined by the American National Standards Institute. Entity relationship diagrams are often used to depict the conceptual data model.
Configuration Accounting	A process for maintaining system baselines, including adding products to a baseline, denoting the components of each product (referred to as configuration items), and monitoring and recording the disposition of requested changes to the system.
Configuration Management	A function which serves to systematically identify the items that make up a system, and formally control any changes or additions to those items, in order to help maintain the integrity of the system, and facilitate communication about the system throughout its life cycle.
Custodianship	The functions and responsibilities of an organization, such as an ADP organization, with physical custody of data that supports the work of another organization, such as a program office. For example, the custodian ensures the physical integrity of the data and software under its control; safeguards the media storing data; ensures the data is secure from unauthorized access, change or destruction; makes data accessible to users; and implements requested hardware or software changes.
Data	Representations of facts, concepts, or instructions in symbols suitable for communication, interpretation or processing by human or automated means.
Data Administration	The management function responsible for the planning, definition, organization, protection, and efficiency of data and data bases within OSWER. The goal of Data Administration is the cost-effective provision of data of sufficient quality to support the OSWER mission.

Data Administration Program	A management initiative which includes policies, standards, and procedures that increase an agency's knowledge and management of the: composition of data, source of data, processing of data, meaning of data, flow of data, and dissemination of data. A successful Data Administration program will improve the management of data by introducing procedures that address: data standards, data requirements determination, data definition, data acquisition or collection, data processing, data storage, data usage (including sharing and access), and data disposal.
Data Attribute	A characteristic of a unit of data such as length, value, or method of representation.
Data Base	A collection of interrelated data stored together with controlled redundancy to serve one or more systems or applications.
Data Base Management System (DBMS)	A software system facilitating the creation and maintenance of a data base and the execution of computer programs using the data base.
Data Collection	The recording and capturing of data on behalf of an organization.
Data Definer	The person or organization who determines the essential qualities or meaning of data, and who prescribes and defines procedures which aggregate and refine data. This includes describing the formatting of the resulting information to serve a specific decision-making context.
Data Dictionary	A centralized repository of information about data, including its meaning, relationship to other data, origin, usage and format. A FIPS data dictionary standard is expected to be approved late in 1988. Called an Information Resource Directory System (IRDS), this standard specifies the capabilities that should be offered for future data dictionaries.
Data Element	The smallest unit of data that has meaning in describing information. A piece of data which would not be meaningful if decomposed further.
Data Entity	A person, place, thing, concept, or event about which an organization may store data. Data entities are to nouns as data elements are to adjectives. That is to say that data entities are the objects being described by data elements. Entity is a synonym for data entity.
Data Flow	A depiction of the movement of information (data) between processes. Data flows are a component of data flow diagrams used by systems analysts to analyze application system requirements.

Data Independence	The property of a data base management system that enables data to be processed independently of access mode, storage method or arrangement. Data independence reduces the need to modify application programs when data storage and access methods are modified.
Data Integrity	The quality of data that exists as long as accidental or malicious destruction, alteration or loss of data are prevented. This results in preservation of data in its intended format, length and contents while within a data base.
Data Life Cycle	The data life cycle begins with the definition of data to support new regulations or other program needs, and includes strategic data planning, data standardization, and the methods and standards during the collection, storing, accessing, and archiving of data.
Data Security	The protection of data against unauthorized disclosure, transfer, modification or destruction, whether accidental or intentional.
Data Steward	See Stewardship.
Data Store	A component of a data flow diagram that represents data at rest or in storage.
Decision Paper	A decision document presented to management. It summarizes those aspects of the analysis and decisions of a given phase or stage that are important to program management, and requests approval to continue the project. The OSWER life cycle model provides for Decision Papers to be prepared at the end of Concept, Definition, Design, Development, and Implementation. The Mission Element Needs Statement (MENS) also is considered a Decision Paper.
Definition	The first stage of the Definition and Design phase. Its purpose is to define specific, detailed functional and data requirements for the system within the context of the System Concept.
Definition and Design	The third phase of the system life cycle, consisting of two stages: Definition and Design. (See individual definitions of each of these terms).
Design	The second stage of the Definition and Design phase. Its purpose is to produce detailed specifications for the system to meet the functional and data requirements within the context of the System Concept.
Development	The first stage of the Development and Implementation phase. Its purpose is to produce a system which is ready for acceptance testing and suitable for implementation.

Development and Implementation	The fourth phase of the life cycle. Its purpose is to produce a complete system, fully tested and available for use in normal production mode. There are two stages in this phase: Development and Implementation.
Domain	A set of all values that a particular data element may posses in actual or potential usage.
Enhancement	A modification to a system that results in substantially improved capabilities and, in some way, alters the structure of the system. Other modifications that do not alter the structure, are referred to as changes. Examples of enhancements include the addition of new data elements, changing the system (or a part of the system) to run in a different software environment, and replacing data entry screens to improve the user interface and/or improve performance.
Entity	A person, place, thing, concept, or event that is of interest to an enterprise. An entity is something about which we store data. Examples of entities are: waste site, contract, employee. An entity has various attributes, or data elements, which should be recorded. Examples for the entity contract could include contract-number, date, and obligation-ceiling. Entity is a synonym for data entity.
Entity Relationship Diagram	A diagram that depicts entities, their key attributes (data elements), and the relationships between entities. The entity relationship diagram is the most commonly used data modeling technique used today. Sometimes system developers refer to entity relationship diagrams as ER diagrams or static data models.
Evaluation	The second stage of the Operation phase. Its purpose is to determine whether the system is effectively meeting the stated requirements, is operating efficiently and is effectively managed.
Implementation	The second stage of the Development and Implementation phase. Its purpose is to produce a fully tested system containing the data needed at start-up, and to provide needed training to the intended users.
Information	Any set of data which has been aggregated by processing in order to establish a specific meaning and serve in a decision-making context.
Information Flow	A depiction of the flow of information between processes. An information flow is another name for a high-level data flow.
Initiation	The first phase of the system life cycle. Its purpose is to define an information management problem within OSWER and to determine whether resources should be committed to exploring ways to address it.

Initiation Decision Paper	A brief document identifying and describing an information management problem. This document is prepared during the Initiation phase.
Life Cycle	See 'System Life Cycle"
Life Cycle Management	The process of managing a system through its life cycle. As practiced by OSWER, it is not a rigid process, but rather a disciplined means for selecting and practicing the management approaches and techniques that are most appropriate for a given information management problem and/or system.
Logical Data Model	A depiction of the logical, or programmatic, data needed to support an organizational mission. The components of a logical data model include data entities and relations; data elements or attributes; keys; secondary keys; and relationships, if data entities are used. The logical data model is a more detailed depiction of the conceptual data model of an organization. It may correspond to the external schema as defined by the American National Standards Institute.
Maintenance	The set of activities that keep systems and data bases in operating condition. Maintenance also focuses on optimizing the performance of existing systems and data bases, without affecting functionality or the structure of the systems or data bases.
Metadata	Data about data, such as its definition or its physical characteristics.
Normalization	The process of reducing a logical data model (structure) to its most basic form, so that the data model is stable, flexible, and without redundancy. A normalized data model is composed of normalized data entities. A normalized data entity includes no repeating groups or data elements among its attributes, contains attributes (data elements) only about the entity being described, and does not include attributes which are dependent upon the key of another entity.
Operation	The fifth phase of the life cycle. Its purpose is to operate the system in normal production mode, monitoring and maintaining its performance, until the end of the life cycle, and then to terminate operation. There are three stages in this phase: Production, Evaluation, and Archive.
Phase	The major segments of the system life cycle. There are five phases in the OSWER system life cycle: Initiation; Concept; Definition and Design; Development and Implementation; and Operation.
Privacy	The right of individuals or organizations to constrain the collection and use of data about themselves.

Production	The first stage of the Operation phase. Its purpose is to make the system available to users, and make required changes and improvements to ensure that it contains to address the information management problem in a cost effective manner.
Project	An organized effort to solve an information management problem. In most cases, a project extends over the entire system life cycle. In some cases a project extends only through the portion of the life cycle that can be foreseen with confidence, e.g., through Production if the timing for ceasing operation is uncertain.
Project Execution	The set of activities which produce the concept, definition, design, and production versions of a system.
Project Management	The set of activities which monitor and control project execution to ensure that they are performed effectively and in accordance with applicable policies, guidances and practices; and that its products solve the identified information management need.
Quality Assurance	A function that ensures that all products of the life cycle are substantively accurate and address the stated information management problem. Quality assurance is accomplished through the efforts or skilled professionals on the project team, and through formal reviews.
Record	A group of related data elements treated as a unit by an application program.
Review	A formal examination of a life cycle product to verify that the system as represented solves the specified information management problem. The OSWER life cycle model provides for five reviews: Concept, Definition, Design, Development, and Post- Implementation. Any review may be repeated as necessary to ensure that all deficiencies have been fully and adequately addressed in the designated products.
Shared Data	Data stored that is created, accessed, updated, or deleted by more than one organizational unit.
Stage	The segments of the system life cycle that occur within certain larger phases. The OSWER system life cycle divides the Definition and Design phase into two stages: Definition and Design. The development phase is divided into two stages: Development and Implementation. The Operation phase is divided into three stages: Production, Evaluation, and Archive. The phases Initiation and Concept are not divided into stages.

Stewardship	The functions and responsibilities of an organizational entity that exercises control over data on behalf of OSWER. Organizations that require data to be collected, processed, stored or used in support of OSWER's mission have stewardship responsibilities. These responsibilities include ensuring that: (1) Only data relevant to OSWER's missions is collected. (2) Data that is collected is of sufficient quality to support OSWER's missions. (3) Data is reused wherever appropriate within OSWER. (4) Data is clearly defined and documented in compliance with established directives. (5) Systems practices under the organization's stewardship conform to EPA Data Administration guidance.
System	An organized set of functions, data, procedures, hardware, software, communications, and/or documentation which enables OSWER to solve a specific information management problem. A system need not necessarily be automated; but most instances of life cycle management will apply to automated information systems.
System Concept	A high-level complete description of a system (including data, processing capabilities, hardware, software and communications). It is produced during the Concept phase and serves as both a check on the validity and completeness of the problem, and the basis for defining more detailed functional and data requirements.
System Life Cycle	The evolution of a system from the initial identification of an information management problem through system termination or replacement.
Threshold Analysis	The process of determining the appropriate review and approval levels for an OSWER system project.
Walkthrough	A highly-structured meeting to review the completeness and quality of selected module(s) of the system, or of the entire system. Walkthroughs are usually conducted by the project team, often are attended by user representatives, and may be held at any point in the system life cycle.

8.2 Additional Reference Materials

8.2.1 Data Modeling References

- Entity Modeling, Ronald Ross, Database Research Group, Boston, 1987.
- Guide on Logical Database Design, Elizabeth Fong, et al., National Bureau of Standards Publication 500-122, 1985.

8.2.2 Physical Data Base Design References

- Computer Data Base Organization, James Martin, Prentice-Hall, Englewood Cliffs, New Jersey, 1975.
- Data Base: Structured Techniques for Design, Performance, and Management, S. Atre, A. Wiley Series, New York, 1980.
- Design and Strategy For Distributed Data Processing, James Martin, Prentice-Hall, Englewood Cliffs, New Jersey, 1981.

8.2.3 Data Documentation References

- Data Dictionary/Directory Systems, Belkis Leong-Hong and Bernard Plagman, John Wiley & Sons, New York, 1982.
- Data Dictionaries and Data Administration, Ronald Ross, AMACOM, New York, 1981.
- Guide on Data Entity Naming Conventions, Judith Newton, National Bureau of Standards Publication 500-149, 1987.